Prediction and learning in the dynamics of speaking

Citation for published version:
Pickering, M & Gambi, C 2015, 'Prediction and learning in the dynamics of speaking', *Language, Cognition and Neuroscience*, vol. 31, no. 4, pp. 514-516. https://doi.org/10.1080/23273798.2015.1117645

Digital Object Identifier (DOI):
10.1080/23273798.2015.1117645

Published In:
Language, Cognition and Neuroscience

Prediction and learning in the dynamics of speaking

Martin J. Pickering
University of Edinburgh

Chiara Gambi
University of Edinburgh

Address for correspondence:
Martin Pickering
Department of Psychology
University of Edinburgh
7 George Square
Edinburgh, EH8 9JZ, U.K.
martin.pickering@ed.ac.uk
Phone: +44 (0)131 650 3447
Fax: +44 (0)131 650 3461

Strijkers and Costa (in press) provide an extensive and detailed analysis of the dynamics of speaking from a behavioural and neuroscientific perspective. They acknowledge that much evidence supports a serial, sequential model in which different components of word production (lexical-semantic components, lexical phonology, motor phonology, and articulation) are constructed in order and in different cortical areas (e.g., lexical-semantic components at 200ms in MTG). Such data are compatible with Indefrey and Levelt's (2004) neurocognitive model, and with psycholinguistic models that are either primarily feed-forward (following Levelt, 1989) or admit some feedback but remain broadly staged (e.g., Dell, 1986). But they then point to many types of data that appear incompatible with such architectures, with evidence suggesting that radically different components (e.g., phonology, semantics) might be activated in parallel or even reordered. How might this come about?

In an important footnote (4), Strijkers and Costa (in press) point out that they focus on cell-assembly models. They note that dual-stream models (Hickok, 2012, 2014; Pickering & Garrod, 2013) may provide an alternative explanation for the data, but do not consider such accounts further. Here, we propose that much of the inconsistent evidence appears highly compatible with such dual-stream models. For example, such models can explain the evidence that suggests both early effects of phonology (during the time-window traditionally ascribed to lexical-semantic processing) and late effects of phonology, by arguing that production involves rapid prediction and slower implementation.

Consider Pickering and Garrod's (2013) account of language production. The speaker forms a communicative intention (production command), for example to name a picture. The production command is sent to the production implementer, and it triggers the retrieval of a set of production representations (semantics, syntax, and phonology) that are constructed during preparation for speech. For example, if a speaker forms the intention to name a kite, production processes would cause the retrieval of the corresponding concept (KITE), lemma (*kite*) and phonological form (/kaIt/). Importantly, these representations are constructed sequentially over several hundred milliseconds (just as Indefrey & Levelt, 2004, assume).

But in addition, an efference copy of the production command is sent to a forward model, which maps from the production command to predicted representations (i.e., those representations the speaker predicts are about to be retrieved as a consequence of executing that production command). To return to our example of a speaker intending to name the picture of a kite, a forward model of this process could compute a prediction of aspects of the semantics (it's a flyable object), of the syntax (it's a noun), and the phonology (it starts with a consonant), *before* using the implementer to construct the corresponding representations. Therefore, predicted representations are typically ready before implemented representations.

Importantly, rapid activation in areas that are not traditionally interpreted as part of the process of implementation has been taken as evidence for computation of forward models. For example, Tian and Poeppel (2010) reported activation of bilateral temporal areas within 170ms during imagined articulation of a syllable. Importantly, they registered spatially similar activation while participants passively listened to the same syllable, and therefore attributed this activation to sensory auditory processing (see also Tian & Poeppel, 2013 for more direct evidence that articulation imagery affects responses in auditory cortex). The fact that activation occurs so rapidly in an auditory area during imagined articulation, when no auditory stimulus is present, suggests that forward model predictions are involved. Such activation is particularly striking because it implies that the speaker has represented the predicted experience of hearing the syllable (which is not part of the production process itself). The pattern of activation reported by Tian and Poeppel (2010) does not mean that Indefrey and Levelt's (2004) map is wrong; the early activation in auditory areas need not reflect a stage of the production implementer. Rather, it may reflect a predicted representation that can be compared with the implemented representation when this is ready.

It is possible to construct a related interpretation of studies such as Miozzo, Pulvermueller, and Hauk (2014), who found evidence for parallel activation of lexical-semantic and phonological-articulatory processing within 200ms. Strijkers and Costa (in press) propose that such findings show that Indefrey and Levelt's (2004) map is incomplete, and that phonology might (under some conditions) become available alongside lexico-semantic information. But it is also possible that the early evidence of phonological activation in fact constitutes another neural manifestation of a forward model. Note that early phonological activation in Miozzo et al's study occurred in the posterior MTG, which has previously been implicated in word form encoding (e.g., Graves, Grabowski, Mehta, & Gordon, 2007), but also in the mapping between lexico-semantic, lexico-syntactic and phonetic/acoustic processing (Gow, 2012; Hickok & Poeppel, 2007; see also Menenti, Segaert, & Hagoort, 2012). Mapping between representational levels in a potentially non-sequential manner is one important feature of forward models (Pickering & Garrod, 2013), which further distinguishes forward models from strictly sequential implemented representations. Our interpretation is therefore compatible with the extensive evidence (reviewed on pp. 21-22) for "a sequential and local component to language production", and is at least a reasonable alternative to an account in which the stages of production are radically revised, as proposed by Miozzo et al. and Strijkers and Costa.

There is another respect in which the evidence of early phonological processing may not be incompatible with Indefrey and Levelt (2004). Using ERPs, Strijkers, Costa, and Thierry (2010) observed a cognate effect within 200ms – the same time-window as the frequency effect. In other words, Spanish-Catalan bilingual speakers showed different responses for words that had overlapping sounds in the two languages (e.g., *libro-llibre*; "book") than for ones that did not (e.g., *mesa-taula*; "table"). Strijkers and Costa (in press) assume that these results reflect near-simultaneous retrieval of lexical (i.e., semantic) and phonological information. But another possibility is that cognates acquire their special status during learning. For example, every time a Spanish learner of Catalan intends to name a book in Catalan, she not only activates the Catalan word *llibre* but also its Spanish cognate *libro* because of the phonological relationship between them. Over time, this leads to a slow increase in the resting activation of *llibre*. (It is possible that this process of relating *llibre* and *libro* may take place "off-line", after production of *llibre.*). With increasing levels of fluency, the speaker ceases to activate *libro* when selecting *llibre*, but the raised activation level remains. In the fluent L2 speaker of Catalan, selecting *llibre* is like selecting a higher-frequency non-cognate, but the speaker need not access phonology early (or indeed activate any aspects of the non-target language; see Costa, Pannunzi, Deco, & Pickering, 2015 for a related proposal). Indeed, Strijkers et al. (2010) offer precisely this explanation for the early onset of the cognate effect when discussing their own results (p. 925).

If such an explanation is on the right lines, cognate status (i.e., a phonological property) becomes implicitly incorporated into the lexical-semantic representation as a consequence of learning. In turn this means that the ease with which the lexico-semantic representation of an L2 word is accessed during production will depend on whether it is a cognate or not. Hence, brain responses should be modulated by cognate status during the early (lexical) stage of word production, as Strijkers et al. (2010) showed. In sum, this account can explain why cognates led to the same P2-range effect as a lexical frequency manipulation in Strijkers et al. (2010) without drastically revising the stages of production.

Rather more speculatively, we can relate such incorporated representations to the use of forward models. Forward models play a major role in fast online control of speech (Guenther, Ghosh, & Tourville, 2006; Hickok, 2012) because they implement learned mappings between production commands and lexico-semantic, motor, and sensory goals. In other words, forward models are the product of learning. For example, a Spanish speaker intending to name a book would have learned through repeated experience that when the concept BOOK is activated, activation of the phoneme /l/ usually follows. This means that when she accesses the concept BOOK she can predict, via a forward model, activation of the phoneme /l/. If this speaker begins learning Catalan, she will repeatedly experience that when the concept BOOK is activated in a Catalan context, activation of the phonologies of both *llibre* and *libro* usually follows. In time, she could predict, via a forward model, aspects of both phonological representations, or perhaps she could just predict the initial sound /l/ very strongly because it is present in both word forms. Such predictions of phonology may lead to patterns of brain activation in which phonological representations appear to be activated in a temporally and spatially non-traditional manner. Importantly, such activation need not reflect the implemented process of activating phonology – something which may still take place in a traditional, largely sequential manner, as assumed by Indefrey and Levelt (2004).

These proposals do not of course invalidate the concerns that Strijkers and Costa (in press) raise with the traditional model. It is highly striking that activation does not always follow the functional and neural model that conceives of speaking as a largely linear process moving from intention to articulation. But they do suggest that Indefrey and Levelt (2004) may characterise a core "implemented" process that is not strongly affected by context. Instead, it is augmented by auxiliary mechanisms (see Pickering & Clark, 2014) whose role is, perhaps, to expedite the difficult act of speaking.

References

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2015). *Do bilinguals automatically activate their native language when they are not using it?* Manuscript submitted for publication. University of Edinburgh.

Dell, G. S. (1986). A spreading-activaction theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321.

Gow, D. W. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain and Language, 121*(3), 273-288.

Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience, 19*(4), 617-631.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280-301.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135-145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*(1), 2-20.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Menenti, L., Segaert, K., & Hagoort, P. (2012). The neuronal infrastructure of speaking. *Brain and Language, 122*(2), 71-80.

Miozzo, M., Pulvermüller, F., & Hauk, O. (2014). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex, 25*(10), 3343-3355.

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences, 18*(9), 451-456.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329-392. doi:http://dx.doi.org/10.1017/S0140525X12001495

Strijkers, K., & Costa, A. (in press). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition, and Neuroscience*.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20*(4), 912-928.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology, 1*, doi: 10.3389/ fpsyg.2010.00166.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience, 25*(7), 1020-1036.